DOCUMENT RESUME

ED 450 710 IR 020 591

AUTHOR Hope, Barbara; Mattingly, Maribeth; Spear, Eric; Glasser, Mike
TITLE Meta Data Integration: Maximize the Potential of "Data about Data."
PUB DATE 1999-00-00
NOTE 14p.; In: EDUCAUSE '99: Celebrating New Beginnings. [Proceedings] (Long Beach, CA, October 26-29, 1999); see IR 020 580. Figures may not reproduce adequately.
AVAILABLE FROM For full text: http://www.educause.edu/conference/e99/proceedings.html.
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Copyrights; Fair Use (Copyrights); Higher Education; Information Networks; Information Policy; Information Systems; Intellectual Property; Internet; *Metadata; Policy Formation

ABSTRACT

The University of Maryland, College Park recognized in the 1990s that its institutional data was an asset that needed to be managed. In this Information Age, data must be turned into knowledge quickly and accurately. The University of Maryland implemented a campus data warehouse, and along with it a comprehensive meta data platform for helping individuals understand the meaning and context of the data they were accessing. With limited resources it was apparent that meta data needed to be captured at a single point of entry and it needed to be available or delivered to multiple points of distribution. This paper gives an overview of how the University of Maryland Meta Data Manipulator works and how it allows for the meta data to be integrated with the data warehouse structures, and the tool used to query the data. Fourteen figures are provided at the end of the paper. (Author/AEF)

Reproductions supplied by EDRS are the best that can be made
from the original document.

This paper is the intellectual property of the author(s). It was presented at EDUCAUSE '99, an EDUCAUSE conference, and is part of that conference's online proceedings. See http://www.educause.edu/copyright.html for additional copyright information.

Meta Data Integration:
Maximize the Potential of 'data about data'

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

C.J. Keller

Barbara Hope, Data Administrator

Maribeth Mattingly, Assistant Data Administrator

Eric Spear, Data Base Administrator, Institutional Research

Mike Glasser, Data Base Administrator, Operations and Enterprise Applications

University of Maryland
College Park, Maryland

ED 450 710

Abstract

The University of Maryland, College Park recognized in the 1990's that it's institutional data was an asset that needed to be managed. In this Information Age data must be turned into knowledge quickly and accurately. The University of Maryland implemented a campus data warehouse and along with it a comprehensive meta data platform for helping individuals understand the meaning and context of the data they were accessing. With limited resources it was apparent that meta data needed to be captured at a single point of entry and it needed to be available or delivered to multiple points of distribution.

This session gives an overview of how the University of Maryland Meta Data Manipulator works and how it allows for the meta data to be integrated with the data warehouse structures and the tool used to query the data.

Introduction

As data are widely distributed throughout an organization, it is critical to understand the meaning of the data and the context in which they are presented. The University of Maryland's Office of Data Administration (ODA) was charged with identifying institutional data elements, defining them, indicating who had responsibility for them and educating the campus community in the use of these institutional data elements. ODA designed a single source meta data application which enabled the office to leverage the comprehensive "data about data" and make it readily available, in multiple ways, to the users. A web-based application, using a Java applet for easy browser access, enables the office to catalog standardized data elements, their definitions, examples, supplementary definitions, keywords, data subsets, transaction system data, and associated code sets. Via the application, data are entered into Oracle relational tables and leveraged with other systems. Data definitions and codes sets are accessed real-time when customers query the University of Maryland Data Warehouse (DW) using the campus client server tool. The same meta data are also accessible to non-warehouse users via a data definition search tool on a campus web site.

By utilizing features in Brio Technology's query products, ODA is able to make meta data available to users as they write queries, increasing their ability to understand the data and more accurately construct a query. Part of ODA's mission is to educate the campus regarding the meaning and use of institutional data elements. ODA believes the education process is best served by having one point of entry and cataloging of meta data and multiple methods of distribution.

This complete integration of meta data allows for the delivery of meaningful information to all types of users, greatly enhances the ability to educate users, provides a contextual reference for querying, and increases the overall friendliness of the warehouse. In addition, it has created an encyclopedia of

IR020591

knowledge for the organization. Because information is an asset of the organization, it needs to be managed and made available throughout the organization.

Background

The University of Maryland (UM) is the flagship campus of the University System of Maryland (USM). As the comprehensive public research university for the State of Maryland and the original 1862 land grant institution in Maryland, UM has the responsibility within the USM for serving as the state's primary center for graduate study and research, advancing knowledge through research, providing high-quality undergraduate instruction across a broad spectrum of academic disciplines, and extending service to all regions of the state. It has a current Carnegie Classification of Research Universities I. The University is located in College Park, Maryland, five miles from Washington D.C. There are 24,454 undergraduate students, 8,257 graduate students, 315 campus departments, 13,136 permanent faculty, staff, and graduate assistants, and 4,000 hourly student employees.

The Office of Data Administration (ODA) was created in 1996 from a CQI effort that recognized that institutional data was an asset that needed to be managed. The Office reports to Operations and Enterprise Applications, one of three major subunits of the Office of Information Technology. It consists of two FTEs that manage the data administration function and the UM Data Warehouse. ODA's mission is to manage the institutional data of the University of Maryland to provide reliable, accurate, secure, accessible data to meet the strategic and management needs of all levels of the campus. The DW is one mechanism for meeting ODA's mission and providing a platform through which the campus meta data "encyclopedia" of data knowledge is distributed.

The first movement towards data warehousing at the University of Maryland, was a proof-of-concept project. An operational system existed to support faculty appointments on campus; however, no effective query mechanism was available. The DW began as a grass roots effort to prove that the DW concept was viable. The success was immediate and the project began to expand, into the personnel arena, contract and grants, student systems, payroll, financial accounting, and budget.

Data for the UM Data Warehouse are extracted from IBM 3090 and HP 9000 transaction databases (DataCom/DB and Image) via in-house programming. Data are loaded into Oracle 8.0 databases on AIX and Unix servers. Campus users with Windows, Mac and UNIX desktops access the data via SQLNET over the campus TCP/IP network.

The University of Maryland DW architecture began with a comprehensive "atomic" level infrastructure from which data marts were built. ODA felt that if it could provide all of the data from the institutional operational systems, then the building blocks would be in place to move up the pyramid to create data marts, data views, and an executive information system. This approach takes longer to implement, but ODA has been very satisfied with this decision. Because a full data subset is brought into the "atomic" level all at once, queries can be run against the subset and joined to other existing subsets. This results in an immediate "win" for the users. It allows for incremental deployment of subsystems rather than waiting for the entire complement of campus data subsets to be added. *Figure 1*

As the "atomic" level data infrastructure was made available via the DW, a method for educating the users in the understanding and use of the data was necessary. ODA provides the campus community with query tool training and also requires users to attend a data training class for each data subset to which they have been given access (i.e., registration, payroll, personnel…). In the process of training users, it was apparent to ODA that readily accessible information about the data and its source was necessary. At the same time the DW development was occurring, ODA was researching and cataloging data definitions and source information about institutional data elements. ODA's goal was to integrate this Meta Data Encyclopedia with the DW and other information delivery mechanisms, in order to track attributes about the "atomic" data elements and greatly enhance user understanding and use of the DW data elements. The resulting Meta Data Encyclopedia is available through all levels of the DW data architecture. *Figure 2*

What is Meta Data and Why is it Important?

Meta Data is information about the data that make up the institution's data infrastructure. Meta data, as the University of Maryland defines it, includes element definitions, policies that may have an affect on the data, keywords to aid in web searching, operational system origins, programming logic where applicable, element code values with descriptions, and units responsible for each element. By building a meta data encyclopedia, the Office of Data Administration is managing the data asset of the campus. Meta data catalogs information about the elements that support the institutional systems of the campus. Until the formation of the Meta Data Encyclopedia, much of the institutional data knowledge resided in the heads of a dozen or so long time campus employees. It was acknowledged that turnover of these personnel would result in a loss of institutional history and memory. And so the UM Meta Data Encyclopedia was born. It differs from the traditional data element dictionary in that it contains detailed definitions of data elements and provides contextual references. It is not uncommon for a definition to be several paragraphs long. Cataloging of the institutional elements attributes has the immediate effect of educating the entire campus community and preserving the data for the future. We cannot afford to have the data asset leave the campus as individuals leave the campus. In addition, with business process re-engineering efforts and implementation of fully integrated administrative application systems, understanding of the data and their relationship across processes is crucial.

Integrated Delivery of Meta Data

The collection and recording of meta data is a monumental task. It may take weeks of researching policy, combing through data dictionaries, and interviewing functional experts to collect the information needed. The magnitude of the process dictated that meta data be entered once, and only once, into a single source database. It needed to be available not only to ODA, but to service offices, the Office of Institutional Studies, to users of the data warehouse and to the general campus community. ODA's goal was to make it available to DW users via their query tools and to campus constituents via a web search. The meta data encyclopedia needed to support the entire campus infrastructure.

As we began to catalog meta data, there were a limited number of products available on the market. Those that were available were financially beyond the means of our campus. Meta data were originally kept in WordPerfect files as a stop gap measure until a database application could be developed. ODA developed the data model for what would become the Meta Data Encyclopedia and partnered with a database administrator who on his own time developed what is called the Meta Data Manipulator web application. It has become the cornerstone of our meta data cataloging.

The Meta Data Manipulator was built by constructing Oracle tables to hold the meta data. A Java application was written that provides the Office of Data Administration with a single point of entry for all the various meta data components (short and long definitions, subset relationships, keywords, operational systems origins, supplementary technical definitions, and code value translations).

By storing the meta data in Oracle tables, it is considered one of the DW subsets, and our query tool is used to create meta data reports, just as it can be used to create reports for other DW data. Not only are the data definitions immediately accessible via the tool, the translations for the code values are as well. Gone are the days when users needed to have lists or books of codes and their translations next to their computers for reference. It is now all at their finger tips. For example, the code of "01" in a DW element called Category Status Cd (code) is described as "Faculty, Tenured" in the corresponding descriptive DW element called Category Status. The codes with their translations for these elements are available within the tool while composing the query. For Category Status each and every available code and translation will be listed in a window.

The meta data infrastructure allows us to integrate the meta data via the DW query process, ODA's URL, and via Oracle for ad hoc reference and reporting. Efficiency has been achieved by inputting the meta data only once at the source and distributing it in multiple ways. It is not only available for our DW customers, but it is available for anyone on campus who has a need to know about data. *Figures 3 and 4*

A unique part of the meta data delivery to the campus is through the query tool which our campus chose. Several years ago, the BrioQuery client server query tool, from Brio Technologies, was selected by the

campus as the tool that ODA would support for accessing data in the DW. Although any tool that is SQL compliant can be used, ODA supports, offers training for, and provides helpdesk coverage only for the BrioQuery product line. By using the BrioQuery product, a user takes advantage of the available meta data definitions/remarks and lookup codes with code definitions. The BrioQuery product has a mechanism that allows us to reference our meta data Oracle tables and display the information in them. While using BrioQuery, a user can lookup the data definitions for both tables and elements directly from the Meta Data Encyclopedia. There is also a mechanism that allows us to link our tables containing codes and their translations with the BrioQuery tool. This has meant that while users write their queries, they have the encyclopedia of meta data accessible from within the query tool and readily available. This is a huge leap for ODA in it's effort to support and educate the campus regarding the institutional data.

Meta Data Manipulator Design

The Meta Data Manipulator is a Java applet written with Symantec Visual Cafe for Java 1.12. The user doesn't need SQL*Net installed on their client machine to connect to the Oracle server because the thin JDBC classes provided by Oracle to connect directly from the applet to the Oracle server are used.. The applet uses the Netscape security mechanism to get outside the Java sandbox, so it won't run with Microsoft's Internet Explorer. Since all of the meta data cataloging is done by ODA and they use Netscape this has not been a problem for our organization. It might have been a problem if our organization used a different web browser. Since it is a web based product there are no other platform restrictions. It was important to ODA to have the ability to catalog meta data from any machine without special hardware/software requirements. Using the browser as the means to access the application enabled this functionality.

The Manipulator contains many features that make the cataloging of data easy and efficient. Data elements are displayed in a scrollable, alphabetic list with indicators denoting if an element does not contain a definition, cannot be found in the UM Data Warehouse, or is a table as opposed to a data element. Data definition text boxes utilize cut and paste features from within the application as well as external to the application. A tab feature enables easy movement amongst attribute types within an element, such as definition, examples, source data, and supplementary information. *Figures 5-9* Functional buttons enable features such as create new element, delete element, rename element, save data, keyword assignments, and subset relationships, and exit. Scrollable lists of data subsets can be associated with data elements. These subsets correlate to the University's data management structure and enable the cataloging of data elements associated with a responsible data steward. Because a data element can be used by more than one transaction system, source system data is cataloged by the source system in which it is located. For each data element and its transaction system, the following are recorded: element name in source system, format and length of data element in source system, machine on which source system is located, system name, and the table name in the source system. This feature is extremely beneficial in locating common data elements across campus transaction systems. It facilitates the process of standardizing data across these systems.

Behind the applet is a set of Oracle tables and views that interface with the Meta Data Manipulator. These tables contain the Remarks (short definition for the element), Examples, Supplementary Definitions (long definition for the element), Transaction system origins, and Attributes (keywords, subsets, security sensitivity). These are populated and maintained by ODA using the Meta Data Manipulator.

There are other tables that map the DW elements to their codes and translations so the user can easily get a list of the data element's code values and associated descriptions within the BrioQuery tool. ODA found that limiting queries on codes alone had little meaning for campus functional users. Limiting queries on descriptions introduced erroneous data subsets when descriptions were misspelled or ordered differently. The solution was to display codes and their descriptions at the same time. The Manipulator allows for the establishment of this Lookup mapping. Without this mapping, an element with a code would only show the available codes to the user without meaningful translations. The Lookup feature of the Manipulator *(Figure 10)* uses a table that contains, for each element, a table identification, the value code, and a short and long translation. We load the code tables from our transaction systems to the DW nightly. An example of an entry in the code table that translates the category status (faculty, staff and

student employment categories) element:

Table Id	Code	Short	Long
HRCIVS	01	Fac Tenured	Faculty, Tenured
	02	Fac On-Track	Faculty, On Tenure-Track
	03	Fac NT-Term	Faculty, Non-Tenured, Term Only

Another table that completes the capability to "lookup" the code with its translation allows Data Administration to attach the Table Id to an element. At the same time Data Administration chooses what they would like to display for the user to see in BrioQuery. This might be the code and its short translation or the code and its long translation, or just one of the translations. An example of the mapping of elements and table ids:

Element	Table	Type of Display
CATEGORY_STATUS_CD	HRCIVS	COMBINED (combines the code with the short translation)
CATEGORY_STATUS	HRCIVS	COMBINED LONG (combines the code with the long translation)

Connecting the Meta Data to BrioQuery

BrioQuery provides the mechanism for each connection to indicate where the data element remarks can be retrieved and where the code "lookup"descriptor records can be found. This is a feature of the BrioQuery product that has allowed us to customize the product to fit the campus' needs. The goal is for the presentation and use of the DW to be as friendly and easy as possible. It is not a platform designed for the typical programmer. It is a platform designed for a typical business manager on campus. *Figures 11 and 12.*

Tiered Data Delivery Approach

ODA's ultimate goal has been to provide the mechanisms to meet the different functional needs of our campus users. The "atomic" level is for those individuals who want to learn the data intricacies and "explore" and analyze the data in depth through ad hoc query building. Data marts and a pre-written repository of queries are for those individuals who want information, but do not have the time to invest in learning all of the details of the "atomic" level. These individuals are our "consumers" or "farmers". The final type of individuals who need information are our executives. They want to click and get immediate answers at their desktops. They cannot invest the time to attend query training or data training. They need pre-written queries that deliver answers to business questions and provide trend analysis for decision making and strategic planning. For this functionality, we deliver web reports at the click of a mouse and have been able to provide the front-end to our DW and establish a true executive information system infrastructure. *(Figure 13)* The Brio Technology suite of products has enabled us to provide this tiered data delivery approach while at the same time incorporating the knowledge from our Meta Data Encyclopedia. To further serve the campus community, ODA has developed a data definition web search that enables campus users to search for data elements in the Meta Data Encyclopedia from ODA's web site. A search results in a display of all relevant data elements and the meta data attributes cataloged for each element. *(Figure14 & 15)*

Conclusion

By responding to the charge that ODA provide accessible information to the campus, the office moved forward to find an integrated solution. As data were made accessible via the DW, meta data had to accompany the process. Limited staff resources required that the capture of meta data be streamlined. It had to be entered once but distributed easily to various applications. The java application via the web provided the mechanism for capturing and maintaining the meta data. Brio Technologies' Brio Query product line (client server and web server) has made it possible for the meta data to be integrated into

the query and reporting tools. Last but not least, the campus has provided a web meta data search mechanism for individuals on campus that wish not to use the Brio product, but need to understand the institutional data. In a DM Review article, Michael H. Brackett summed it up very appropriately, "A data resource is the heart of an intelligent, learning, information-driven public or private sector organization. Operational data, historical data, analytical data, predictive data, and meta data are all part of that data resource and must be formally managed and integrated within a common data architecture to provide high-quality, meaningful support to the business." The University of Maryland agrees and has taken steps to fully integrate the meta data into the overall data architecture on the campus. So far it seems that this approach has produced meaningful results for many and is proving to be a correct design.

Meta Data Integration:Maximize the Potential of 'data about data'

Figures

Data Administration, University of Maryland, College Park, MD

Figure 1

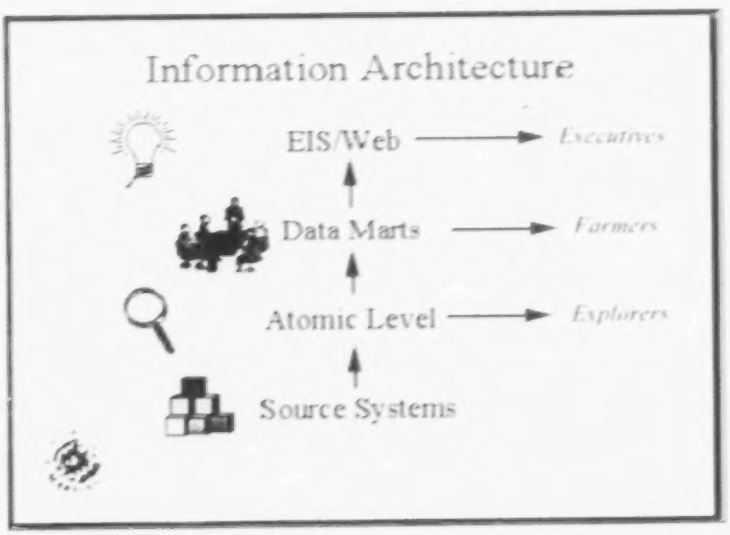


Figure 2

BEST COPY AVAILABLE

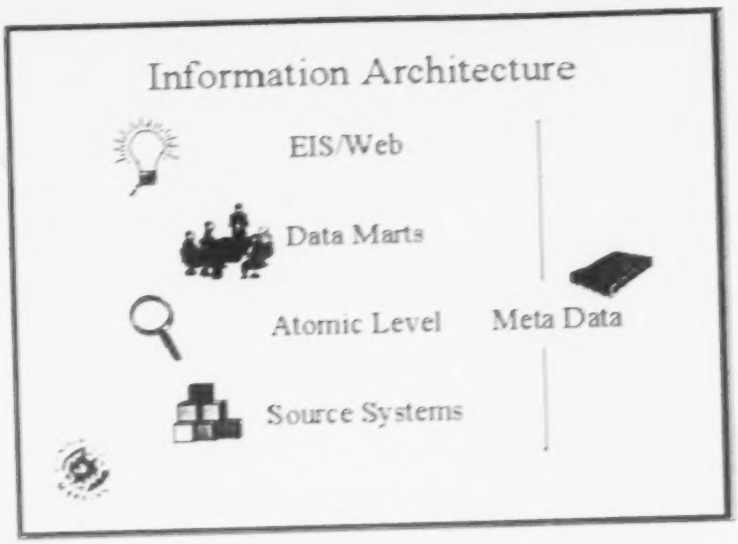


Figure 3

Figure 4

BEST COPY AVAILABLE

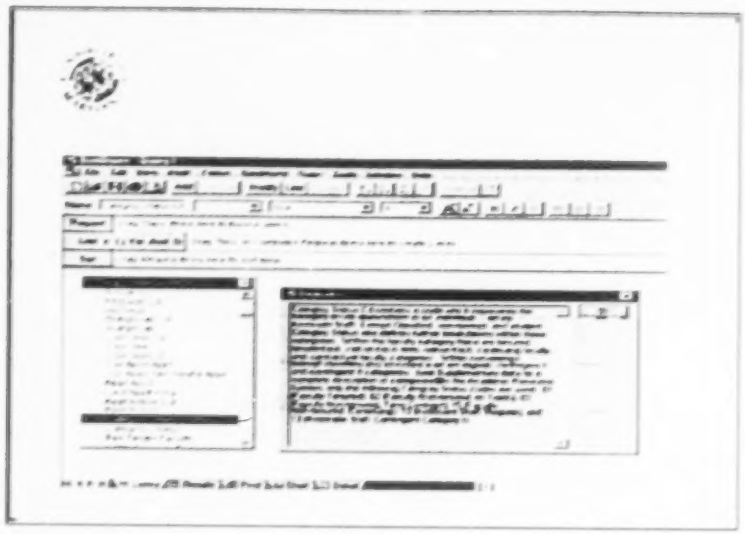


Figure 5

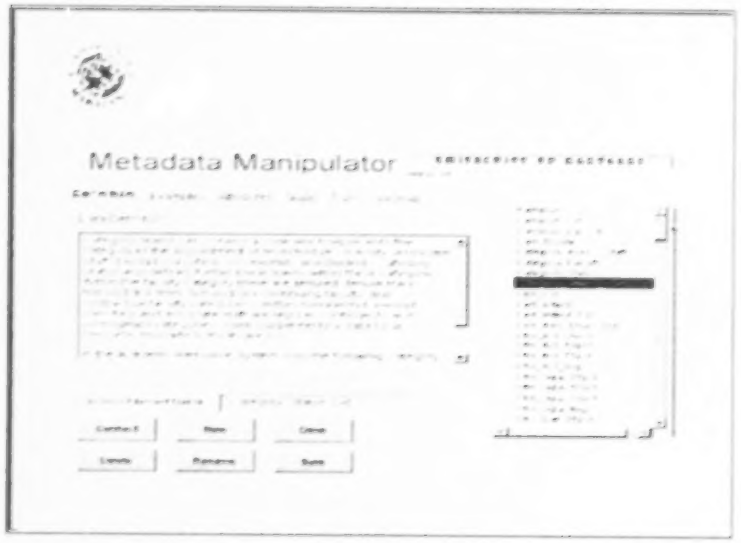


Figure 6

BEST COPY AVAILABLE

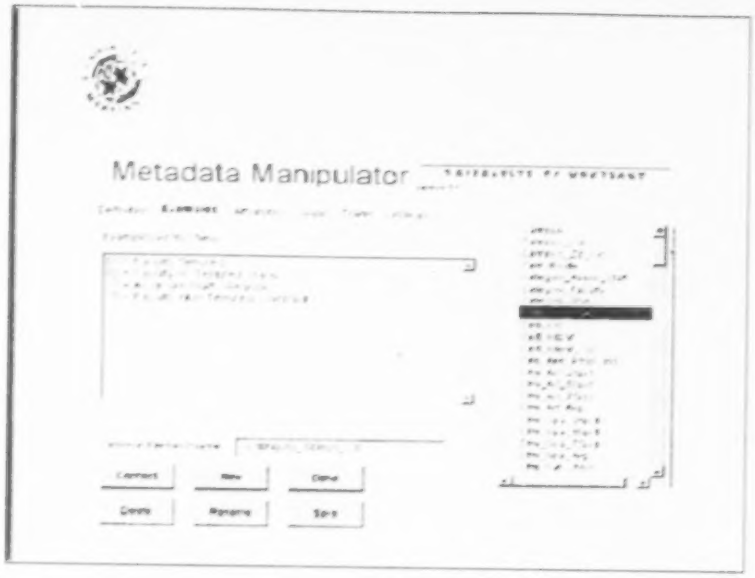


Figure 7

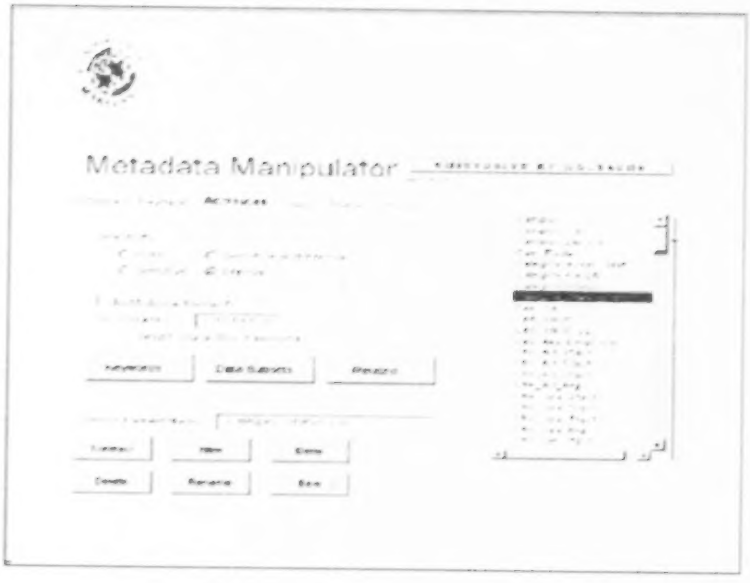


Figure 8

BEST COPY AVAILABLE

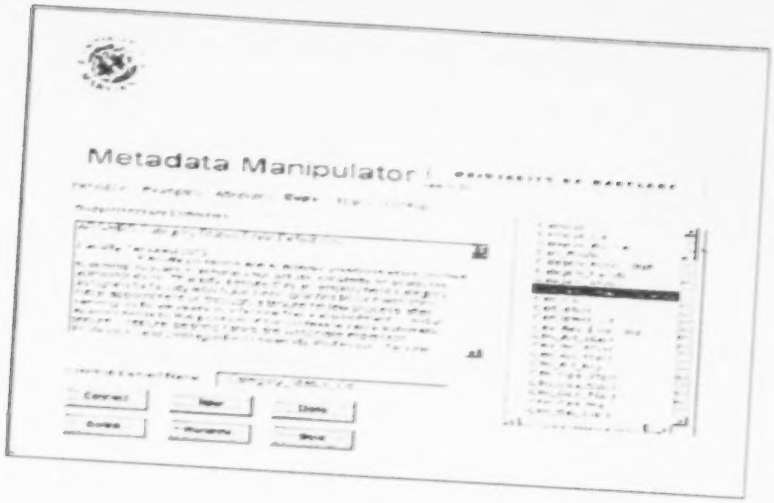


Figure 9

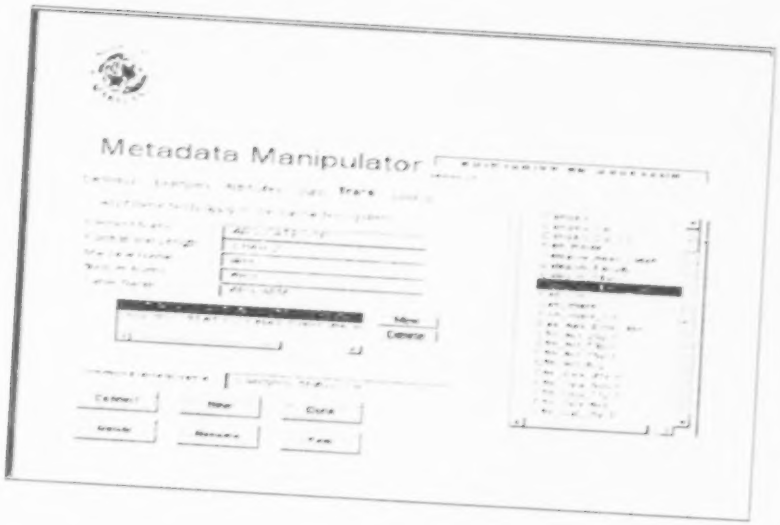


Figure 10

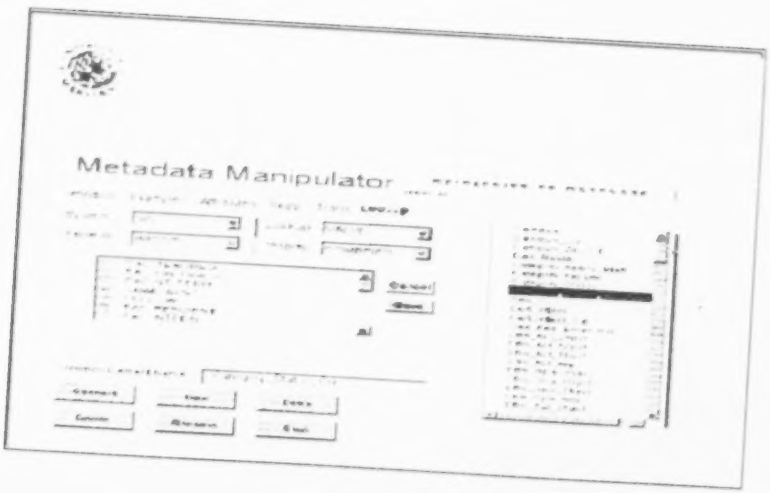


Figure 11

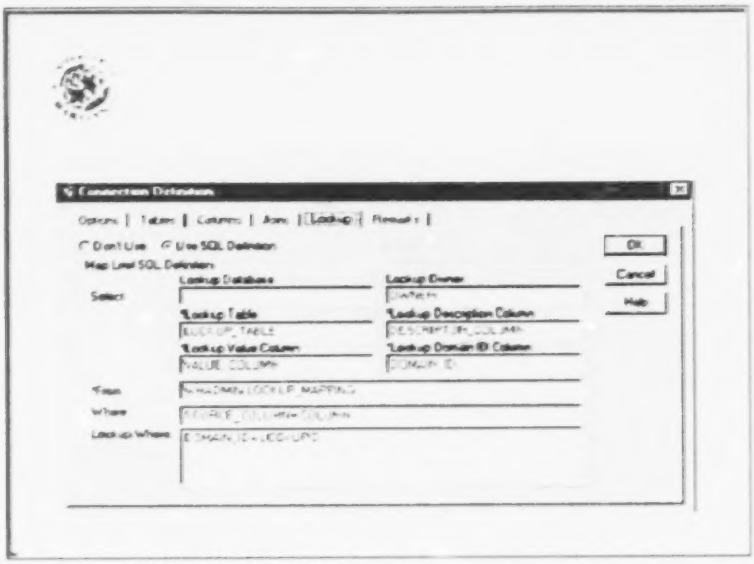


Figure 12

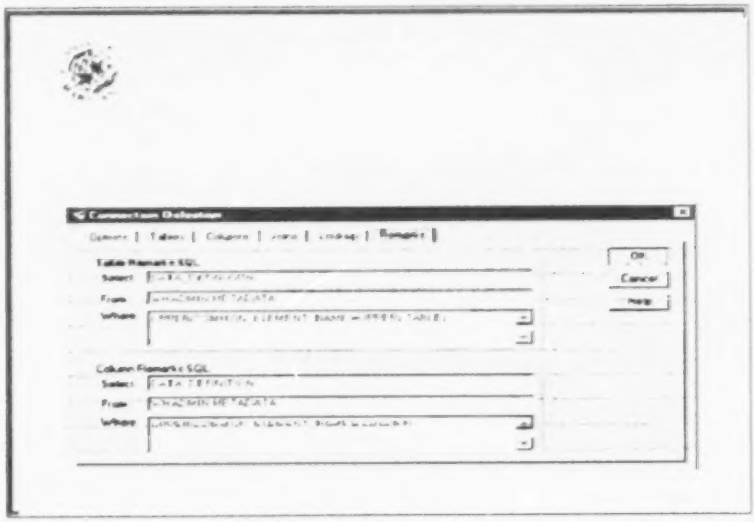


Figure 13

BEST COPY AVAILABLE

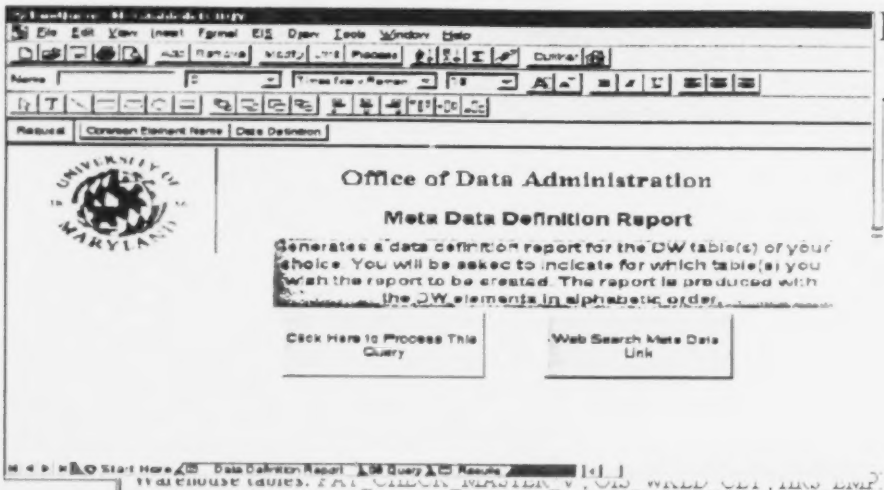
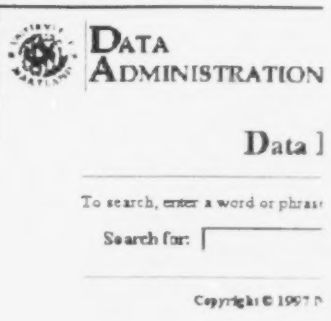

Office of Data Administration
Meta Data Definition Report
Generates a data definition report for the DW table(s) of your choice. You will be asked to indicate for which table(s) you wish the report to be created. The report is produced with the DW elements in alphabetic order.
Click Here to Process This Query
Web Search Meta Data Link

Figure 14



Warehouse tables: PAY_CHECK_MASTER_V ; OIS_WKED_CRT ; HRS_EMP
HRS_EMPLOYEE_GENERAL ; HRS_EMPLOYEE ; COMBINED_EMPLOYEE_
COMBINED_EMPLOYEE_TITLES_V ; ARS_APPOINTMENT_MASTER ; ARS_APPOINTMENT_MASTER_V ;
ARS_CURRENT_APPTS ; BUDGETED_POSITIONS ; HRS_EMPLOYEE_CRS ; HRS_EMPLOYEE_PHYSPL ;
HRS_EMPLOYEE_V ; PAY_CHECK_MASTER

Institutional Element: N

Definition:

Category Status Cd contains a code which represents the category of the appointment of an individual - Faculty, Associate Staff, Exempt Classified, non-exempt, and student. Category Status also defines further breakdowns within these categories

BEST COPY AVAILABLE

EDUCAUSE *Information Resources Library*

Transforming Education Through Information Technologies

Abstract

Category: Papers Presented at EDUCAUSE annual conferences

ID Number:	EDU9902
Title:	Meta Data Integration: Maximize the Potential of 'Data about Data'
Author:	Barbara Hope, Maribeth Mattingly, Eric Spear, Mike Glasser
Organization:	University of Maryland
Year:	1999
Abstract:	The University of Maryland, College Park recognized in the 1990's that it's institutional data was an asset that needed to be managed. In this Information Age data must be turned into knowledge quickly and accurately. The University of Maryland implemented a campus data warehouse and along with it a comprehensive meta data platform for helping individuals understand the meaning and context of the data they were accessing. With limited resources it was apparent that meta data needed to be captured at a single point of entry and it needed to be available or delivered to multiple points of distribution. This paper gives an overview of how the University of Maryland Meta Data Manipulator works and how it allows for the meta data to be integrated with the data warehouse structures and the tool used to query the data.

This material is available in the following electronic formats:

html

Select one of the icons above to retrieve the material in that format. We also have definitions and instructions for setting up your computer to download these formats.